DOCUMENT RESUME

ED 076 612 TM 002 593

AUTHOR Rice, James A.
TITLE Benton's Visual Retention Test: New Age, Scale Score and Percentile Norms for Children.
PUB DATE 72
NOTE 6p.; Paper presented at American Psychological Association, 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Elementary Grades; Evaluation Techniques; *Memory; *Norms; *Perceptual Motor Coordination; Psychological Tests; Technical Reports; *Test Interpretation; Test Reliability; *Visual Perception

ABSTRACT

The Benton Visual Retention Test which is designed to assess visual perceptual, visual motor, and visuoconstructive abilities can give school personnel greater precision and range in testing. The standardization of this instrument was tested on 700 Houston elementary school students. Chronological age differences were maintained and correlation variances of IQs were slight. The test allows for interpretation of individual differences such as memory deficiency and perceptual deficiency. (RS)

FORM 510

PRINTED IN U.S.A.

FILMED FROM BEST AVAILABLE COPY

APA Paper 1972

ED 076612

TM 002 593

Benton's Visual Retention Test: New Age, Scale Score, and Percentile Norms for Children

James A. Rice

University of Houston

Tests such as the Bender Gestalt, Draw-A-Person, and Visual Retention Tests have long been used to assess aspects of perceptual-motor functioning. These have become, despite a lack of educational validity, prominent vehicles for school psychologists, often determined to infer such diagnoses as minimal brain dysfunction on the basis of a marginal performance on a test for which the norms, if they exist, are also marginal.

Koppitz made a notable contribution in 1963 with the publication of children's norms for the Bender Gestalt Test. This tended to objectify interpretation, but validation studies have in general failed to produce promising results. The Bender Gestalt and Visual Retention tests are clinical instruments which are most frequently used in cutoff fashion. Samples of ostensibly normal and deviant subjects are examined. A statement may subsequently be made about an individual as to the probability of his being deviant. Unfortunately, such instruments resist standardization due to the ceilings that are invariably pressed in the course of attempting to utilize them beyond the effective range. It is one thing to make a probability statement about an individual, e.g., he cannot do the task and the odds are 7 in 9 that he is deviant; it is quite another to impose a quantitative range on performance in a heavily skewed distribution, so typical of age distributions in the upper reaches on tests of perceptual-motor functioning.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEH REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

If one is willing to restrict the range of applicability, then a reasonable standardization is possible. Consider, for example, the Bender Gestalt or Visual Retention Tests. Originally designed for use with adults, these tests span a number of years for children during which performance will improve as a function of age, until a level is reached that corresponds to adult competency. Within that span, the test may be properly subjected to standardization; beyond that span, it may not.

Koppitz' children's norms for the Bender lose applicability in the eighth year, the standard deviations of the age distributions exceeding the means, the distributions obviously being skewed as the result of a ceiling effect. Judgments are tenuous beyond this range. Second, the Bender is a copying task, yet considerable research has demonstrated that memory dysfunction is among the more prominent correlates of brain trauma, and would thus appear related to what we have been calling minimal brain dysfunction, if, indeed, the syndrome has educational utility.

The present investigation was directed toward providing school personnel with an instrument of greater precision and range. I refer to the Visual Retention Test (Benton, 1963), published by the Psychological Corporation.

In the manual, the test is defined as "...a clinical and research instrument designed to assess visual perceptual, visual motor and visuo-constructive abilities." It is frequently administered as a memory task, Adminstration A, in which each of 10 designs are exposed for ten seconds, then immediately reproduced from memory by the subject; or, as a copying task, Administration C, or both. The score differential between the two administrations, when sizeable, may be presumed to have diagnostic significance. Three sets of 10 designs each, Forms C, D, and E, constitute the

test stimuli, and two, C and D, are considered in this study. Scoring may be either the number of designs correctly reproduced, or the number of errors.

While the general purpose of the present study was standardization, five specific objectives were set.

(1) Present VRT norms for children are age scores only, and do not extend below eight years. However, a number of recent studies have demonstrated applicability of the VRT down to the age of five years. Thus, a downward extension to the approximate age of five was the first objective.

(2) There is suggestive evidence that Forms C and D are not of equal difficulty level. This issue was to be resolved in the study, and if supported, separate normative distributions would be developed.

(3) Age, scale and percentile distributions would be developed for a normative group (IQ 85 - 115), tested within six weeks of a point 8 months beyond their birthdates, kindergarten through grade six, and no repeaters. The 8th month was selected for a purpose which was not relevant to the present studies.

(4) To identify ceilings beyond which normative interpretation is tenuous on the basis of a clinically meaningful difference between adjacent age distributions.

(5) To determine test-retest reliability coefficients for two of the grade distributions, kindergarten and grade three.

Procedures

A sample of more than 700 children from the Aldine Independent School District, suburban to Houston and with a school population of over 25,000, were given the Visual Retention Test, both copy and memory administrations,

and the Otis-Lennon Mental Ability Test. Of these, 521 were in regular classes and 389 had IQ's between 85 and 115. These comprised the normative sample. Approximately equal numbers of boys and girls were examined, and within the conditions of age earlier discussed, selection was random.

The examinations were given by first and second year graduate students in psychology while enrolled in a course in psychological appraisal. All received a thorough introduction to the administration and scoring of the VRT prior to field testing, and interjudge agreement in practice scoring sessions was found to be quite satisfactory.

Results

Chronological age differences were maintained at almost exactly 12-month intervals, at mean age levels 5-8 through 11-8. No IQ differences were found among the grade levels, nor were significant sex differences observed.

Form C is more difficult than Form D under the copying condition. For number-correct scoring, the difference maintains throughout the effective range of the test, as it does for number-of-errors scoring.

Form D is more difficult than Form C under the memory condition. For number-correct scoring the difference is greater in the middle age range; and for number-of-errors scoring the difference is substantial at the younger age levels and collapses at the nine-year level.

Correlations of the scores with IQ varied between .13 and .28, and do not appear operationally significant, but these have not been corrected for restriction of range of intelligence.

For both forms and administrations, correlation coefficients between number correct and number of errors vary from -.79 to -.94.

While number correct and errors are significantly correlated for the normative group, the interpretative issue for a single child is somewhat different. For example, a child having a number-correct score of five must by definition have an error score of at least five, there being ten designs in all. But it is theoretically possible for another child with a number-correct score of five to have an error score of 15. This approach to interpretation demonstrates the successful quantification of a clinical or qualitative observation--that is, the generally poorer performance of the second child, and does so without the invocation of judgment on the part of an examiner, however well trained. It provides an opportunity to avoid loose clinical interpretation in the presence of a reasonable and objective alternative (See fourth case on sample interpretation sheet).

Age score distributions were derived by fitting second degree curves to the score means at each age level for every condition of administration, scoring and form, a total of eight. The obtained values were plotted, curves fitted, and age scores corresponding to every possible raw score value read back from the resulting graph.

Scale score distributions with a mean of 100 and standard deviation of 15, and percentile distributions, were computed for all conditions. These distributions constitute the norms. No distribution was prepared where raw score mean differences between two adjacent age levels were less than .5 (the clinically significant difference essential to alter a diagnosis taken by Benton in the manual to be one). This is the criterion used to avoid the trap of over-interpretation, that is, forcing the ceiling.

While the captions on the norms give the applicable age in round numbers, i.e., 5, 6, 7, etc., the age means for which these distributions are applicable are at the eighth month, i.e., 5-8, 6-8, 7-8, etc. Thus,

age ranges for the distributions are 5-2 through 6-1, 6-2 through 7-1, 7-2 through 8-1, etc. However, considering the dearth of convincing validity data for tests of this nature, the use of the norms as presented is very probably adequate for most purposes.

Test-retest reliability coefficients are reported on the reverse side of the last page of the norms in the handout, Tables 18 and 19. When corrected for restriction of range of intelligence, these coefficients vary from .85 to .91, and may be considered acceptable for an essentially clinical instrument which is judgmentally scored. Test-retest interval was approximately two months.

Finally, the separate handout presents four possible interpretative situations, all with different diagnostic outcomes. The first is of an average child in his fifth year; the second, an illustration of a memory deficiency; third, a likely perceptual deficiency where, in absolute terms, the child copies the designs as well as he can reproduce them from memory; fourth, an example of the qualitative inference noted earlier, with one child performing much less well than the other; and fifth, a severe memory deficit, in which case one is well advised to inquire as to intellectual functioning. If it parallels prowess in copying, the memory deficit may have far more serious educational implications.